DOCUMENT RESUME

ED 197 664 HE 013 438

AUTHOR Marsh, Herbert W.
TITLE Students' Evaluations of Tertiary Instruction: Testing the Applicability of American Surveys in an Australian Setting.
PUB DATE 80
NOTE 23p.; Tables with factor analysis results may not reproduce well.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Comparative Education; *Cross Cultural Studies; Factor Analysis; *Faculty Evaluation; Foreign Countries; Higher Education; *Measurement Techniques; *Rating Scales; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; Teaching Skills
IDENTIFIERS *Australia; United States

ABSTRACT

Two student evaluation surveys developed in the United States were administered to a sample of University of Sydney students to determine their applicability in Australia. The Endeavor Instructional Rating Form (Frey et al.) measures seven components of effective teaching that have been demonstrated with the use of factor analysis in several different settings, and ratings on this form have correlated with student learning. The SEEQ survey (Students' Evaluation of Educational Quality) by Marsh has been factor analyzed and nine factors have been identified (learning/value, enthusiasm, organization/clarity, group interaction, rapport, breadth of coverage, examinations/grading, assignments, workload/difficulty). Students were asked to evaluate one of the best and one of the worst lecturers in their university experience, to indicate inappropriate items, and to select the most important items. Each of the 63 items was seen as appropriate by most students, each item was chosen by at least a few students as being most important, and all items, except those related to workload/difficulty, differentiated between best and worst lecturers. Separate factor analyses of the two surveys revealed the same factors that had been identified in American settings. Furthermore, there was good agreement between the factors from the two surveys that were hypothesized to measure the same components of effective teaching. It is concluded that the findings demonstrate the feasibility of evaluating effective teaching in Australian universities and the appropriateness of two American surveys to an Australian setting. A list of the factors in each survey, the results of oblique factor analysis, and references are included. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED197664

Students' Evaluations of Tertiary Instruction: Testing the Applicability of American Surveys in an Australian Setting

Herbert W. Marsh

University of Sydney, Australia

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Herbert W. Marsh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE013438

Running Head: Students' Evaluations

Grateful acknowledgements are extended to the students in my Education III research course who helped me conduct this study, and to Dr. Ian D. Smith who offered helpful comments during the course of the study. Requests for reprints should be sent to Herbert W. Marsh, Department of Education, University of Sydney, Sydney, N.S.W. 2006, Australia.

Abstract

Two student evaluation surveys developed in the United States were administered to a sample of Australian university students. Students were asked to evaluate one of the best and one of the worst lecturers in their university experience, to indicate inappropriate items, and to select the most important items. Each of the 63 items was seen as appropriate by most students, each item was chosen by at least a few students as being most important, and all items -- except those related to Workload/Difficulty -- differentiated between best and worst lecturers. Separate factor analyses of the two surveys revealed the same factors that had been identified in American settings. Furthermore, there was good agreement between the factors from the two surveys that were hypothesized to measure the same components of effective teaching. The findings demonstrate the feasibility of evaluating effective teaching in Australian universities and the appropriateness of two American surveys to an Australian setting.

Students' Evaluations of Tertiary Instruction: Testing the Applicability of American Surveys in an Australian Setting

Students' evaluations of teaching effectiveness are commonly collected at American institutions of higher education, and are widely endorsed by both students and faculty (Centra, 1979; Astin & Lee, 1966; Seldin, 1975). The purposes of these evaluations are variously to provide: 1) a source of diagnostic feedback to faculty about the effectiveness of their teaching; 2) a measure of teaching effectiveness to be used in tenure/promotion decisions; and, 3) a source of information for students to use in the selection of courses and lecturers. While the first purpose is nearly universal, the second two are not. At some universities student input is required before faculty are even considered for promotion, while at others the inclusion of students' evaluations in such deliberations is optional. Similarly, the results of student ratings are published and sold in campus bookstores at some universities, while at others the results are considered to be strictly confidential.

The use of students' evaluations, especially for tenure/promotion decisions, has not been without opposition, and investigation of different aspects of the ratings has stimulated considerable research. Particularly in the last few years this has become one of the most frequently studied areas in American educational research. In contrast to the wide use of student ratings in the United States and Canada, they apparently have not been systematically collected in either the United Kingdom or Australia (Smith, 1980). Furthermore, there has been no attempt to empirically test student ratings or the applicability of rating instruments developed in the United States. The purpose of this study is to describe two such instruments, and to report upon an investigation of their applicability in an Australian setting.

The two evaluation instruments used in this study are the SEEQ survey (Students' Evaluations of Educational Quality) developed by the author (see Marsh, 1980b; in press) and the Endeavor Instructional Rating Form devised by Frey (Frey, 1978; Frey, Leonard & Beatty, 1975). While other instruments have been developed (see Centra, 1980 for a summary), both these multifactor instruments were explicitly developed to measure separate components of effective teaching, and both have been shown to be reliable, valid, and little affected by a variety of sources of potential bias.

The Endeavor form measures seven components of effective teaching that have been demonstrated with the use of factor analysis in several different settings (Frey, Leonard & Beatty, 1975). Frey has shown that ratings on Endeavor are correlated with student learning (Frey, 1973; 1978; Frey, Leonard & Beatty, 1975). In these studies, as well as similar studies conducted with SEEQ, student ratings are collected in large multisection courses. Different sections of the same course are taught by different instructors, but each section is taught according to a similar course outline, has similar goals and objectives, and most importantly is tested with the same standardized final examination. Those sections that rate teaching to be most effective near the end of term are also the sections that perform best on the final examination. Frey (1978) has also argued that "pedagogical skill", as measured with the Endeavor form, is relatively unaffected by class size, expected course grade, and instructor rank, but is correlated with research productivity. However, ratings of "rapport" (class discussion and availability for individual attention) were more strongly affected by class size and expected grades, but were not related to research productivity.

SEEQ and the research that led to its development has been recently summarized (Marsh, 1980b). Factor analysis has identifed the nine SEEQ factors in several different studies (Marsh, 1978; in press). Furthermore, factor

analysis of faculty self evaluations of their own teaching effectiveness with the SEEQ form also indentified these same factors (Marsh, in press; also see Marsh, Overall & Kesler, 1979). Various indicies of reliability, including intraclass correlations and coefficient alphas, all indicate that the reliability of these factors is generally at least .9 (Marsh, 1980b). Additionally, when the same students were asked to reevaluate teaching effectiveness several years after graduation from their university programs, the average correlations -- based on ratings of 100 classes -- was .83 (Overall & Marsh, 1979a). Ratings on SEEQ have been successfully validated against the ratings of former students (Marsh, 1977), student learning as measured by objective examination (Marsh, Fleiner & Thomas, 1975; Marsh & Overall, 1980), affective course consequences such as the application of the course materials and plans to pursue the subject further (Marsh & Overall, 1980), and faculty self evaluations of their own teaching effectiveness (Marsh, Overall & Kesler, 1979; Marsh, in press). None of a set of 16 "potential biases" (e.g., class size, expected grade, and prior subject interest) could account for more than 5% of the variance in SEEQ ratings (Marsh, 1980a), and many of the relationships were inconsistent with a simple bias explanation (e.g., more difficult courses that require more time outside of class were rated more favorably). Finally, feedback from SEEQ, particularly when coupled with a candid discussion with an external consultant, led to improved ratings and better student learning (Overall & Marsh, 1979).

In the present investigation items from both SEEQ and the Endeavor form were administered to a broad sample of Australian tertiary/university students. Students were asked to evaluate the teaching effectiveness of a best and a worse lecturer, to indicate inappropriate items, and to select the most important items. These criteria, in addition to factor analysis of the ratings, were used to test the applicability of these American instruments in an Australian setting.

Methodology

Sample

The evaluation survey was administered to a total of 158 students currently enrolled at the University of Sydney. Approximately 30% of these students were enrolled in a course on Human Growth and Development, an introductory level course offered in the Faculty of Arts. The remainder were recruited on an ad hoc basis in various campus libraries, the student union, and departmental lounges. Students were asked to participate, given a brief description of the study, guaranteed of the confidentiality of their responses, and ensured that they would not be asked to identify themselves in any way. Students, if they agreed to participate, were given a self-addressed, stamped envelope with the survey, and asked to return it within three weeks. Approximately half of the surveys actually were returned, and represented a broad cross section of university students. Students were asked to indicate the department in which they would receive their degree, and a total of 25 different departments were named. The courses that were actually evaluated came from 30 different departments.

Survey

The evaluation survey consisted of three pages. The first page consisted of instructions and demographic items, and requested that students select one of the best and one of the worst lecturers that they had experienced at the University of Sydney. Students were asked to limit their choices to "lecturers who were primarily responsible for an instructional sequence that lasted at least one term". Students were then instructed to fill out two separate questionnaires, one for the best lecturer and one for the worst - each containing 63 evaluation items. Items were to be answered on a five-point response scale that varied between "1-Very Poor or Very Low or Almost Never" and "5-Very Good or Very High or Almost Always". An additional "not appropriate" response

categorey was provided for items that were not relevant to the course being evaluated (items left blank were also counted as "not appropriate" responses). After completing the ratings for a given lecturer, students were asked to select up to five questions that they "felt were most important in describing either positive or negative aspects of your overall learning experience in this instructional sequence."

Statistical Analysis

Each of the evaluation items was initially tested in terms of: 1) its ability to discriminate between best and worst lecturers; 2) its appropriateness (i.e., the lack of "not appropriate" responses); 3) its importance (i.e., the number of "most important" nominations); and, 4) how well it correlated with other items designed to measure the same component of effective teaching. Separate factor analyses were performed on the SEEQ and Endeavor forms. Factor scores derived from these factor analyses were correlated to determine the relationship between SEEQ and Endeavor factors. Finally a set of best items -- best in terms of factor analysis, discrimination, appropriateness, and importance -- was selected from the entire set of items.

Results and Discussion

Evaluation of Individual Items

Preliminary inspection of the SEEQ and Endeavor surveys revealed considerable overlap in the factors defined by each. Five SEEQ factors (Learning/Value, Group Interaction, Individual Rapport, Examinations, and Workload/Difficulty) correspond closely to five Endeavor factors (Student Accomplishments, Class Discussion, Personal Attention, Grading, and Workload). A sixth SEEQ factor, Organisation/Clarity, has been divided into two factors on the Endeavor form (Presentation Clarity and Organisation/Planning). Three SEEQ factors, Instructor

Enthusiasm, Breadth of Coverage, and Assignments/Readings, do not correspond to any of the Endeavor factors. On the basis of this preliminary inspection the set of 34 SEEQ items and 21 Endeavor items were divided into 10 content areas that were hypothesized to correspond to 10 components of effective teaching (see Table 1). In addition to these 55 items, eight additional items were also classified into one of the 10 categories to make a total of 63 items.

Students evaluated a best and a worst lecturer with the entire set of 63 items. The best lecturers were evaluated more favorably ($p < .001$) on all items except those in the Workload/Difficulty factor (See Table 1). The differences were largest for the Learning/Value/Accomplishment, Instructor Enthusiasm, and Presentation Clarity factors. While there was little difference between best and worst lecturers in terms of the difficulty of courses which they teach, courses taught by the best lecturer were judged as more difficult on five or six items (one to a statistically significant extent). These findings demonstrate that both the SEEQ and the Endeavor items are able to clearly discriminate between lecturers that Australian students have selected as being best and worst.

Insert Table 1 About Here

Students were specifically asked to indicate items that were inappropriate for evaluating one of their lecturers or the course he/she taught. Only two of the 63 items were judged to be inappropriate (including those that left an item blank) by more than 10% of the students. (see Table 1) One of these concerned the availability of personal attention (some Australian students did not seek it), and the other asked about feedback from examinations (some courses did not administer exams or administered only a final examination that was not returned to students). In general, items falling into the Group Interaction/Discussion, Individual Rapport/Personal Attention, and Grading/Examination factors were most frequently seen as inappropriate. However, every single item was judged to be appropriate by at least 80% of the students, and most were appropriate for 95%

or more of them. These results indicate that the SEEQ and Endeavor items are appropriate in an Australian setting.

Students selected as many as five items that they felt were most important in describing either positive or negative aspects of the overall learning experience. Each of the 63 items (see Table 1) received at least three "most important" nominations, and at least one item in each of the 10 categories received 17 or more nominations. The three most frequently selected items were "teaching style held your interest" "presented clearly and summarized", and "was enthusiastic about teaching". Items falling into the Instructor Enthusiasm and Presentation Clarity factors were nominated most frequently. While some of the items and some of the factors appear to be more important, the nominations were spread widely over the entire set of items. This suggests that each item measures a potentially important component of effective teaching.

For each item, the average correlation between that item and other items in the same factor, and between that item and all other items was computed (see Table 1). Items that do not correlate with any other items are probably not related to effective teaching in an Australian setting. Items that correlate as highly with all items as with items designed to measure the same component of effective teaching lack specificity, though they may measure some generalized notion of effective teaching. Inspection of Table 1 indicates that every item is substantially correlated with other items that measure the same factor. However, most of the items also have substantial correlations with the entire set of items, and a few are more highly correlated with the entire set of items than with items specifically designed to measure the same component. This suggests that there is a large general factor underlying the student ratings, and that the different factors are highly correlated. While this large general factor or halo effect is undesirable, the manner in which the data was collected may have produced this occurence. Students were specifically asked to select their "best" and "worst" lecturers, suggesting that most lecturers would be

judged as "generally good" or "generally bad". A smaller general factor might be expected if students had also considered lecturers who fell between the two extremes. Nevertheless, the fact that most items were more highly correlated with other items within the same factor suggests that the separate factors are meaningful. The test of this tentative appraisal will be the results of the factor analyses described below.

Factor Analysis of the SEEQ and Endeavor Surveys

The items from Marsh's SEEQ survey (those marked "M" on Table 1) were factor analyzed (see Table 2). Marsh (1978; in press) has demonstrated that the survey measures nine separate components of effective teaching in American settings. Factor analysis of the ratings by Australian students clearly identify eight of these factors. Except for two items from the Examinations/Grading factor, each item loaded substantially on the factor it was designed to measure and had smaller loadings on each of the other factors. Only one of the Examination items loaded substantially on that factor. These results show that at least eight of the nine SEEQ factors are well defined in an Australian setting.

Insert Table 2 About Here

A separate factor analysis of the 21 items from Frey's Endeavor survey (those marked "F" in Table 1) was also performed. The results of this analysis clearly identifies each of the seven Endeavor factors (see Table 3). With one exception (item F13), each item loads substantially on the factor it was designed to measure and has smaller loading on each of the other factors. Frey's Grading factor, unlike the Examination factor from SEEQ, appears to be well defined. The Endeavor factor emphasizes the determination of grades rather than the actual examinations. The Endeavor factors that are most highly correlated, Clarity and Organization/Planning, are the two factors that are combined into a single factor on the SEEQ form. While these factors are highly correlated, the correlation is

less than the reliability of either (coefficeint alpha -- see Table 4), thus supporting their separation. These findings demonstrate that all seven of the Endeavor factors are well defined in an Australian setting.

Insert Table 3 About Here

Factor scores, weighted averages of the standardized ratings for each item (Nie, et al., 1975), were computed for the nine SEEQ factors and the seven Endeavor factors. Correlations among these different factor scores are presented in Table 4. This set of correlations is somewhat analagous to a multitrait-multimethod matrix in which the different factors correspond to the multiple traits and the different surveys are the multiple methods. Criterion developed by Campbell & Fiske (1959; also see Marsh & Hocevar, 1980) can be applied to test for convergent and discriminant validity. Convergent validity refers to the extent of agreement between the two different surveys.

Discriminant validity refers to the distinctiveness of the different factors; specifically, do the highest correlations appear between the SEEQ and Endeavor factors that were designed to measure the same component of effective teaching.

Insert Table 4 About Here

The application of the multitrait-multimethod analysis is somewhat complicated by the fact that the SEEQ and Endeavor surveys were not specifically designed to measure the same components of effective teaching.

However, the four criteria proposed by Campbell and Fiske can be adapted. They are:

1) Agreement on SEEQ and Endeavor factors that match should be substantial (a criterion of convergent validity)
2) Correlations between these matching factors should be higher than the

correlations among different SEEQ factors or among different Endeavor factors (a criterion of discriminant validity)

3) Correlations between these matching factors should be higher than correlations between other SEEQ and Endeavor factors that were not designed to measure the same component (a second criterion of discriminant validity).

4) The pattern of correlations among SEEQ factors should be similar to the pattern among Endeavor factors (e.g. if the two SEEQ factors Group Interaction and Individual Rapport are highly correlated, then so should the corresponding Endeavor factors). Satisfaction of this criterion suggests that correlations among different factors represent correlations between the underlying traits being assessed rather than a method or halo effect.

On the basis of the preliminary inspection of the two surveys, each of the seven Endeavor factors was predicted to correspond most closely to one of the SEEQ factors (the Organization/Clarity factor on SEEQ was hypothesized to be a combination of the Organization/Planning and Clarity factors on Endeavor). Each of these seven correlations is substantial and statistically significant (median r = .80), thus supporting their convergent validity. Six of these correlations -- the one exception being the correlation between the Grading (SEEQ) and Examinations (Endeavor) factors -- are higher than other correlations in the multitrait-multimethod matrix, supporting the discriminant validity of the factors.

In spite of the good evidence for both convergent and discriminant validity, correlations among the different factors are larger than desirable. Among the SEEQ factors, correlations between Learning/Value, Instructor Enthusiasm, and Organization/Clarity factors are all over .50, as is the correlation between Group Interaction and Individual Rapport. Among the Endeavor factors Organization/Planning and Clarity are correlated .67, while correlations between Personal

Attention and Group Discussion, Organization/Planning and Student Accomplishments, and Clarity and Student Accomplishments are all about .6. Several points are relevant to interpreting these high correlations. First, as previously mentioned, part of these high correlations might be a function of the manner in which the data was collected. Secondly, each of these correlations is substantially less than the reliabilities of the factors. Third, there was a similarity in the pattern of correlations among SEEQ factors and among Endeavor factors (Campbell & Fiske's fourth criterion -- also see Marsh & Hocevar, 1980). This final point suggests that the underlying dimensions being measured may in fact be correlated in a way that is independent of the method of data collection. Nevertheless, there is still the possibility of a substantial method or halo effect that should be explored in future research.

In summary, SEEQ and Endeavor factors that have been demonstrated in American settings were also clearly identifed in the evaluations by Australian students. The multitrait-meltimethod analysis of the two surveys offered good evidence for both their convergent and discriminant validity. This argues for the multidimensionality of student ratings, the distinctiveness of the different factors measured by SEEQ and Endeavor, and the generality of these factors beyond the specific items included on each instrument.

A Combination of the Two Surveys

A set of 43 items was selected to best define the 10 hypothesized factors. Items were selected primarily on the basis of factor analysis, but criteria such as discrimination between best and worst lecturers, appropriateness, and most important nominations were also considered. Factor analysis (see Table 5) provides good support for the 10 factors that were hypothesized. Every item loads at least .35 on the factor it was hypothesized to measure (most were over.5), and no item loaded higher than .35 on any other factor (most were under.20). The clarity of this factor analysis offers further support for multidimensionality of student ratings, and the generality of the evaluation factors as measured by the

SEEQ and Endeavor instruments.

Insert Table 5 About Here

Summary and Implications

Items from two student evaluation surveys developed in the United States were administered to a sample of Australian university students. These students were asked to select a best and worst lecturer from their university experience, to evaluate both with the combined survey, and to select the items they felt were most important. Each of the 63 items was seen as appropriate by at least 80% of the students (most by 95% or more), each item was selected by at least a few students as being most important, and all the items -- except those in the Workload/ Difficulty factor -- differentiated well between the best and worst lecturers. Separate factor analyses of the items from the two surveys identified the factors each had been designed to measure. Furthermore, there was good agreement between factors from the two different surveys that were hypothesized to measure the same components of effective teaching. Finally, a factor analysis of items from both surveys offered even clearer support for the hypothesized set of evaluation factors.

These findings clearly demonstrate that teaching effectiveness can be measured in an Australian setting, that evaluation forms developed in the United States are appropriate to an Australian setting, and that the same components that underlie effective teaching in the United States are also relevant in Australia. The findings also suggest that the extensive range of research into student ratings that has been conducted in the United States may also apply to Australian settings. Future research, perhaps employing the surveys used in this study, is needed to determine the extent of the validity of student evaluations of teaching effectiveness in Australian settings, and to document problems involved in the actual implementation of such programs.

References

Astin, A.W. & Lee, C.B.T. Current practices in the evaluation and training of college teachers. Educational Record, 1966, 47, 361-375.

Campbell, D.T. & Fiske, D.W. Convergent and discriminant validiation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105

Centra, J.A. Determining Faculty Effectiveness. San Francisco: Jossey-Bass,1979.

Frey, P.W. A two-dimensional analysis of student ratings of instruction. Research in Higher Education, 1978, 9, 60-91.

Frey, P.W., Leonard, D.W. & Beatty, W.W. Student ratings of instruction: Validation research. American Educational Research Journal, 1975, 12 327-336.

Marsh, H.W. The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal. 1977, 14, 441-447

Marsh, H.W. Students' Evaluations of Instructional Effectiveness: Relationship to Student, Course, and Instructor Characteristics. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March, 1978. (ERIC Document Reproduction Service No. ED 155 217).

Marsh, H.W. The influence of student, course and instructor characteristics on evaluations of university teaching. American Educational Research Journal, 1980, 17, 219-237, (A)

Marsh, H.W. Students' Evaluations of College/University Teaching: A description of Research and an Instrument. Paper presented at the Annual Meeting of the Australian Association for Research in Education, Sydney, Australia, November, 1980. (ERIC Document Reproduction Service No. ED). (B)

Marsh, H.W. & Cooper, T.L. Prior subject interest, students' evaluations, and instructional effectiveness. Multivariate Behavioural Research (in press)

Marsh, H.W. Validity of students' evaluations of college teaching; a multi-trait-multimethod analysis. Journal of Educational Psychology, (in press)

Marsh, H.W., Fleiner, H., & Thomas, C.S. Validity and usefulness of student evaluations of instructional quality. Journal of Educational Psychology, 1975, 67, 833-839

Marsh, H.W. & Hocevar, D. An Application of LISREL Modeling to Multitrait-Multimethod Analysis. Paper presented at the Annual Meeting of the Australian Association for Research in Education, Sydney, Australia, November, 1980. (ERIC Document Reproduction Service No. ED).

Marsh H.W. & Overall, J.U. Validity of students' evaluations of teaching effectiveness: Cognitive and Affective criteria. Journal of Educational Psychology. 1980, 70, 468-475.

Marsh, H.W., Overall, J.U. & Kesler, S.P. Validity of student evaluations of instructional effectiveness: A comparison of faculty self-evaluations and evaluations by their students. Journal of Educational Psychology. 1979, 71, 149-160.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. Statistical Package for the Social Sciences. New York, McGraw-Hill, 1975

Overall, J.U. & Marsh, H.W. Students' evaluations of instruction: A Longitudinal study of their stability. Journal of Educational Psychology. 1980, 72, 321-325 (a)

Overall, J.U. & Marsh, H.W. Mid-term feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology. 1979, 71, 856-865 (b)

Seldin, P. How Colleges Evaluate Professors. Croton-on-Hudson, New York: Blythe-Pennington, 1975.

Smith, I.D. Student assessment of tertiary teachers. Vestes, 1980, 27-32.

Table 1

Hypothesized Factors, Individual Items and Their Characteristics (N= 316 "best" and "worst" lecturers)

Evaluation Items (paraphrased) and Hypothesized Factors	Discrimination: Mean of "Best"	Discrimination: Mean of "Worst	Discrimination: t-value	Number of "Not Appropriate" responses	Number of "Most Important" nominations	Correlation with other items on same scale	Correlation with all other items
LEARNING/VALUE/ACCOMPLISHMENTS							
M1 Course challenging & stimulating	4.16	2.32	17.26***	1	63	.764	.728
M2 Learned something valuable	4.15	2.38	15.51***	2	51	.823	.790
M3 Increased subject interest	3.92	1.94	17.53***	4	30	.846	.807
M4 Learned & understood subject matter	4.08	2.34	17.62***	1	16	.821	.792
M30[a] OVERALL COURSE RATING	4.35	2.20	21.32***	0	32	.852	.828
F19 Understood advanced material	3.98	2.48	12.45***	11	16	.664	.775
F20 Ability to analyze issues	3.96	2.15	17.38***	3	32	.856	.802
F21 Increased knowledge & competence	4.19	2.36	18.97***	2	26	.856	.783
O3 Interest after completion of course	4.17	2.10	20.19***	3	24	.851	.813
INSTRUCTOR ENTHUSIASM							
M5 Enthusiastic about teaching	4.54	2.41	19.76***	1	89	.807	.784
M6 Dynamic and energetic	4.25	1.75	25.20***	1	63	.900	.839
M7 Enhanced presentation with humor	3.99	1.74	18.76***	4	50	.793	.771
M8 Teaching style held your interest	4.49	1.25	45.64***	1	116	.898	.913
M31[a] OVERALL INSTRUCTOR RATING	4.67	1.55	45.15***	1	39	.887	.884
PRESENTATION CLARITY (Organization)[b]							
M9 Lecturer explanations clear	4.41	1.70	32.22***	2	67	.902	.811
M10 Materials well explained & prepared	4.22	1.88	25.10***	4	34	.874	.852
M12 Lectures facilitated taking notes	4.38	1.78	25.67***	2	53	.815	.778
F1 Presentations clarified materials	4.29	1.58	32.68***	0	62	.919	.841
F2 Presented clearly & summarized	4.54	1.77	33.41***	0	90	.927	.847
F3 Made good use of examples	4.24	2.32	18.51***	2	22	.785	.767
PLANNING/OBJECTIVES (Organization)[b]							
M11 Course objectives stated & pursued	4.15	2.15	19.55***	7	30	.789	.804
F13 Presentations planned in advance	4.45	3.11	11.99***	2	42	.672	.695
F14 Provided detailed course schedule	3.83	2.58	9.14***	3	18	.627	.491
F15 Activities orderly scheduled	3.90	2.43	13.28***	28	4	.756	.760
O1 Time distibuted over topics appr'ely	4.14	2.89	11.56***	6	9	.645	.671
O7 Aims & objectives clear to students	4.06	2.10	18.57***	3	21	.805	.772
O8 Announced goals &/or criteria	3.79	2.29	13.54***	5	9	.789	.733
GROUP INTERACTION/DISCUSSION							
M13 Encouraged class discussions	3.50	2.49	9.91***	32(10%)	35	.903	.755
M14 Students shared knowledge/ideas	3.33	1.99	11.21***	25	12	.856	.763
M15 Encouraged questions & gave answers	3.79	1.96	17.39***	7	12	.796	.836
M16 Encouraged expression of ideas	3.44	2.14	11.71***	12	10	.887	.792
F10 Class discussion was welcome	3.65	2.34	10.16***	31	6	.872	.735
F11 Students encouraged to participate	3.36	2.20	9.02***	32(10%)	35	.841	.879
F12 Encouraged students to express ideas	3.37	2.10	11.02***	18	3	.919	.782
INDIVIDUAL RAPPORT/PERSONAL ATTENTION							
M17 Friendly toward individual students	3.91	2.51	12.18***	10	23	.820	.713
M18 Welcomed students seeking help/advice	3.86	2.37	13.34***	7	24	.884	.788
M19 Interested in individual students	3.76	2.03	15.37***	15	10	.857	.775
M20 Accessible to individual students	3.56	2.45	10.02***	21	10	.789	.717
F7 Listened & was willing to help	3.97	2.43	13.70***	21	16	.836	.759
F8 Able to get personal attention	3.94	2.70	10.74***	58(18%)	18	.786	.745
F9 Concerned about student difficulties	3.62	2.14	13.40***	9	21	.849	.798
BREADTH OF COVERAGE							
M21 Contrasted various implications	4.00	2.54	12.68***	16	11	.703	.672
M22 Gave background of ideas/concepts	3.74	2.47	11.91***	10	14	.746	.713
M23 Gave different points of view	3.87	2.65	11.09***	27	22	.716	.727
M24 Discussed current developments	3.81	2.29	13.37***	17	9	.709	.785
GRADING/EXAMINATIONS							
M25 Examination feedback valuable	3.49	2.21	10.64***	41(13%)	8	.551	.612
M26 Evaluation methods fair/appropriate	3.77	2.93	8.06***	22	17	.700	.593
M27 Tested course content as emphasized	4.02	2.97	9.44***	26	13	.626	.612
F16 Grading was fair & impartial	3.97	2.43	13.70***	21	16	.785	.654
F17 Grading reflected student performance	3.60	2.76	7.94***	31	10	.745	.622
F18 Grading indicative of accomplishments	3.60	2.68	9.36***	28	6	.765	.659
O6 Exams covered materials broadly	3.93	2.89	10.01***	23	5	.701	.704
ASSIGNMENTS/READINGS							
M28 Readings/texts were valuable	3.48	2.99	3.81***	12	10	.484	.399
M29 They contributed to understanding	3.75	2.87	7.39***	2	20	.662	.604
O2 They encouraged further exploration	3.41	2.08	11.24***	8	12	.616	.711
O4 Assignment/reading expectations clear	3.74	2.78	8.29***	10	4	.597	.695
O5 They were integrated into the course	3.89	2.89	9.11***	14	11	.670	.700
WORKLOAD/DIFFICULTY							
M32 Course difficulty (Easy-Hard)	3.17	3.38	- 1.90	2	8	.693	-.002
M33 Course workload (Light-Heavy)	3.23	3.22	0.11	0	20	.718	.100
M34 Course pace (Too Slow-Too Fast)	3.14	3.11	0.21	0	23	.513	.020
F4 Students had to work hard	3.46	3.41	0.71	0	12	.739	.114
F5 Course required a lot of time	3.32	3.16	1.31	0	5	.800	.159
F6 Course workload was heavy	3.33	3.03	2.42*	0	5	.811	.300

a--Overall rating items were not specifically designed to measure any one factor but previous research (Marsh, 1980, in press) has shown that these are the factors that they are most related to in American studies.

b--These two factors were combined into one on the Marsh survey but separated on the Frey survey.

Table 2

Factor Analysis of SEEQ items (N= 316 sets of ratings)

Evaluation items (paraphrased)	Factor Pattern Matrix I	II	III	IV	V	VI	VII	VIII	IX
I LEARNING/VALUE									
M1 Course challenging & stimulating	**.57**	.20	.04	.03	-02	.05	.19	.06	.11
M2 Learned something valuable	**.63**	.06	.05	.14	-05	.02	.18	.14	.01
M3 Increased subject interest	**.58**	.12	.15	.04	.04	.18	-.01	.09	.02
M4 Learned/Understood subject matter	**.40**	.07	.34	.11	-.04	.10	.00	.17	-.14
M30 OVERALL COURSE RATING	**.49**	.15	.15	-.02	.09	.13	.16	.10	-.04
II ENTHUSIASM									
M5 Enthusiastic about teaching	.06	**.55**	.10	.11	.09	.11	.10	-.03	-.05
M6 Dynamic & Energetic	.04	**.69**	.08	.07	.06	.08	.19	.01	-.03
M7 Enhanced presentations with humor	.13	**.46**	.02	.15	.12	.07	.17	.01	-.11
M8 Teaching style held your interest	.16	**.44**	.35	.03	.10	.10	.13	-.05	-.01
M31 OVERALL INSTRUCTOR RATING	.23	**.40**	.30	.06	.11	.13	.10	.01	-.03
III ORGANIZATION/CLARITY									
M9 Instructor Explanations Clear	.07	.25	**.56**	.03	.13	.02	.17	-.02	-.08
M10 Course materials prepared & clear	.11	.20	**.57**	.02	.08	.15	.04	.06	.00
M11 Objectives stated & pursued	.28	.06	**.36**	.00	.13	.15	.20	.06	.03
M12 Lectures facilitated note taking	.07	.36	**.44**	.01	.06	.05	.10	.00	.06
IV GROUP INTERACTION									
M13 Encouraged class discussions	.04	.07	.00	**.79**	.08	.05	.02	.10	.01
M14 Students shared ideas/knowledge	.02	.08	.10	**.70**	.08	.06	.07	.06	.01
M15 Encouraged questions & answers	.11	.05	.27	**.36**	.31	.11	.10	-.10	.06
M16 Encouraged expression of ideas	.11	.07	.03	**.66**	.23	.11	.07	-.08	.01
V INDIVIDUAL RAPPORT									
M17 Friendly towards students	.15	.12	-.12	.10	**.67**	.20	.06	-.09	-.03
M18 Welcomed seeking help/advice	.12	.13	-.04	.24	**.55**	.13	.08	-.02	-.01
M19 Interested in individual students	.11	.16	.08	.12	**.58**	.21	-.01	-.05	.01
M20 Accessible to individual students	-.08	.14	-.01	.18	**.49**	.17	.03	.18	.09
VI BREADTH OF COVERAGE									
M21 Contrasted implications	-.04	.07	.09	.02	.03	**.42**	.36	.15	-.06
M22 Gave background of ideas/concepts	-.01	.08	.10	.22	-.13	**.50**	.20	.12	-.07
M23 Gave different points of view	-.06	-.01	.14	.07	.08	**.64**	.11	.13	-.01
M24 Discussed current developments	.14	.13	.08	.19	- 02	**.38**	.15	.14	-.01
VII EXAMINATIONS/GRADING									
M25 Examination feedback valuable	.00	-.01	.16	.16	.25	-.30	**.53**	.30	-.09
M26 Eval. methods fair/appropriate	.04	.07	.14	.11	.15	.04	**.00**	.40	-.05
M27 Tested exphasized course content	.08	-.03	.31	-.05	.19	.12	**.08**	.24	-.02
VIII ASSIGNMENTS									
M28 Reading/texts valuable	.04	-.06	.04	.01	.00	.05	.15	**.50**	.01
M29 Added to course understanding	.15	.21	-.06	.00	.07	.04	-.11	**.66**	.15
IX WORKLOAD/DIFFICULTY									
M32 Course difficulty (easy-hard)	-.15	-.01	.03	.09	-.03	-.02	-.03	.05	**.84**
M33 Course workload (light-heavy)	.14	-.03	.06	-.04	.06	.00	-.01	.05	**.76**
M34 Course pace (too slow-too fast)	.05	.01	-.13	-.09	.09	.05	.39	-.09	**-.43**

NOTE: The oblique factor analysis was performed with the commercially available SPSS routine (Nie, et. al., 1975). Factor loadings in the boxes are the loadings for items designed to measure each factor.

Table 3

Factor Analysis of the Endeavor Items (N=316 sets of ratings)

Evaluation Items (paraphrased)	Factor Pattern Matrix						
I PRESENTATION CLARITY	I	II	III	IV	V	VI	VII
F1 - Presentations clarified materials	.516	.006	.075	.132	.079	.044	.395
F2 - Presently clearly & summarized	.545	-.011	.113	.091	.146	-.018	.344
F3 - Made good use of examples	.452	-.010	.082	.127	.124	.129	.234
II WORKLOAD							
F4 - Students had to work hard	.075	.813	-.029	-.039	-.008	.068	-.038
F5 - Course required a lot of time	-.005	.900	.069	.027	-.004	-.049	.036
f6 - Course workload was heavy	-.058	.833	-.016	.051	.126	.014	.066
III PERSONAL ATTENTION							
F7 - Listened & was willing to help	.061	.031	.709	.115	-.020	.074	.095
F8 - Able to get personal attention	.184	.041	.591	.159	.058	.155	-.095
F9 - Concerned about student difficulties	-.113	.025	.620	.193	.135	.055	.223
IV CLASS DISCUSSION							
F10- Class discussion was welcome	.042	-.025	.245	.690	-.055	.089	.075
F11- Students encouraged to participate	.038	.015	-.003	.894	.045	.042	-.039
F12- Encouraged students to express ideas	-.050	.044	.221	.724	.092	.026	.099
V ORGANIZATION/PLANNING							
F13- Presentations planned in advance	.520	.053	.180	-.073	.319	.039	-.092
F14- Provided detailed course schedule	.047	.033	.006	.027	.578	.006	.100
F15- Activites orderly scheduled	.181	.068	.074	.024	.601	.162	.022
VI GRADING							
F16- Grading was fair & impartial	.161	.020	.077	.050	-.075	.702	.075
F17- Grading reflected student performance	-.075	.011	.080	.044	.163	.844	-.004
F18- Grading indicative of accomplishments	.046	.006	.044	.048	.117	.730	.103
VII STUDENT ACCOMPLISHMENTS							
F19- Understood advanced material	.223	-.278	.100	.046	.222	.082	.351
F20- Ability to analyze issues	.073	.092	.039	.050	.184	.160	.637
F21- Increased knowledge & competence	.058	.055	.092	.028	.141	.079	.696

NOTE: The oblique factor analysis was perfored with the commercially available SPSS routine (Nie, et. al., 1975). Factor loadings in the boxes are the loadings for items designed to measure each factor.

Table 4

A Multitrait-Multimethod Matrix of Correlations Among SEEQ and Endeavor Factors (N=316 sets of ratings)

	SEEQ Evaluation Factors								
SEEQ Evaluation Factors	MI	MII	MIII	MIV	MV	MVI	MVII	MVII	MIX
Learning/Value (MI)	(92)								
Instructor Enthusiasm (MII)	55	(95)							
Organization/Clarity (MIII)	52	60	(93)						
Group Interaction (MIV)	26	39	24	(94)					
Individual Rapport (MV)	31	47	33	54	(93)				
Breadth of Coverage (M VI)	39	49	47	42	40	(88)			
Examinations/Grading (M VII)	46	52	48	33	32	46	(81)		
Assignments/Readings (M VIII)	37	15	35	22	18	33	39	(84)	
Workload/Difficulty (M IX)	06	-04	-15	-05	-03	-01	20	07	(91)

	SEEQ Evaluation Factors									Endeavor Evaluation Factors						
Endeavor Evaluation Factors	MI	MII	MIII	MIV	MV	MVI	MVII	MVIII	MIX	FI	FII	FIII	FIV	FV	FVI	FVII
Presentation Clarity (F I)	47	71	[82]	23	35	49	55	32	-13	(92)						
Workload (F II)	14	06	-02	05	02	05	32	20	[75]	00	(94)					
Personal Attention (F III)	40	56	41	63	[81]	57	43	29	-05	43	04	(90)				
Class Discussion (F IV)	29	45	20	[88]	57	39	33	22	-03	23	05	60	(91)			
Organization/Planning (F V)	58	59	[68]	26	35	56	51	39	06	67	16	43	21	(85)		
Grading (F VI)	39	31	43	28	39	36	[34]	50	-04	31	06	40	25	41	(90)	
Student Accomplishments (F VII)	[80]	63	70	33	37	49	56	39	-10	60	03	44	29	60	35	(85)

Table 5

Factor Analysis of the Combined Set of Items (N = 316 sets or ratings)

Evaluation Items (paraphrased)	I	II	III	IV	V	VI	VII	VIII	IX	X
I LEARNING/VALUE										
M1 Course challenging & stimulating	.56	.11	.15	.07	.04	.03	.03	.03	.10	.13
M2 Learned something valuable	.62	.04	.05	.13	.13	-.01	.05	.09	.09	.04
M3 Increased subject interest	.57	.12	.08	.09	.00	.08	.12	.05	.10	.00
F21 Increased knowledge & competence	.55	.08	.19	.08	.07	.00	.09	.05	.09	-.01
II INSTRUCTOR ENTHUSIASM										
M5 Enthusiastic about teaching	.07	.57	.14	.01	.12	.11	.09	.02	.00	-.01
M6 Dynamic and energetic	.06	.72	.06	.07	.03	.10	.11	.06	.03	.04
M7 Enhanced presention with humor	.11	.36	.13	.09	.13	.19	.10	.03	.03	-.08
M8 Teaching style held your interest	.16	.39	.34	.12	.03	.12	.10	.04	.03	.00
III PRESENTATION CLARITY										
M9 Lecturer explanations were clear	.02	.26	.55	.16	.07	.03	.04	.06	.07	-.05
M10 Materials well explained & prepared	.03	.23	.43	.22	.04	.00	.12	.12	.13	-.02
F1 Presentations clarified materials	.14	.29	.48	.12	.06	.02	.10	.02	.08	-.04
F2 Presented clearly & summarized	.08	.20	.62	.16	.03	.04	.03	-.03	.12	-.03
F3 Made good use of examples	.01	.19	.37	.09	.03	.04	.22	.09	.16	-.03
IV PLANNING/OBJECTIVES										
M11 Course objectives stated & pursued	.21	.08	.24	.36	-.02	.14	.15	.08	.05	.03
F14 Provided detailed course schedule	.01	.06	.00	.72	.01	-.03	.08	-.02	.03	.02
O7 Aims & objectives clear to students	.15	.06	.32	.41	-.03	.08	.07	.12	.05	.00
O8 Announced goals &/or criteria	.07	.00	.17	.58	.05	.11	.04	.11	.04	.03
V GROUP INTERACTION										
M13 Encouraged class discussions	-.02	.05	-.02	.06	.82	.05	.07	.03	.13	.01
M14 Students shared knowledge/ideas	.01	.06	.12	.01	.70	.05	.14	.06	.05	.03
M16 Encouraged expression of ideas	.10	.08	-.02	.06	.62	.27	.12	.00	-.07	.02
F10 Class discussion was welcome	.08	.05	.03	-.03	.70	.18	.00	.09	.05	-.03
VI INDIVIDUAL RAPPORT										
M17 Friendly towards individual students	.08	.13	.07	.03	.18	.57	.06	.02	-.06	-.01
M18 Welcomed seeking help/advice	.08	.10	.04	.07	.23	.64	.04	.03	-.02	-.01
M20 Accessible to individual students	-.12	.09	.00	.07	.14	.53	.13	.13	.13	.07
F7 Listened & was willing to help	.06	.15	.06	.03	.14	.58	.10	.10	-.04	-.05
VII BREADTH OF COVERAGE										
M21 Contrasted various implications	.01	.12	.08	.14	.02	.04	.52	.06	.08	.04
M22 Gave background of ideas/concepts	.04	.08	.04	.09	.16	-.03	.58	.00	.10	-.07
M23 Gave different points of view	.03	.03	.05	.02	.04	.11	.68	.10	.00	.01
M24 Discussed current developments	.16	.10	.05	.10	.08	.10	.47	.00	.10	.01
VIII GRADING										
M26 Evaluation methods fair/appropriate	.01	-.02	.14	-.17	.09	.05	.09	.64	.21	-.02
F16 Grading was fair & impartial	.02	.06	.13	-.02	.05	.04	.09	.70	.09	.00
F17 Grading reflected student performance	.05	.11	-.09	.21	.03	.09	.00	.80	-.02	.01
F18 Grading indicative of accomplishments	.08	.00	-.01	.09	.03	.03	.05	.70	.08	-.02
IX ASSIGNMENTS/READINGS										
M28 Readings/texts were valuable	.03	-.06	-.03	.12	.03	-.08	.17	.06	.42	.03
M29 They contributed to understanding	.10	.02	-.03	-.09	-.05	.17	.05	.12	.74	.08
O2 They encouraged further exploration	.27	.29	-.14	.20	.11	.01	-.11	-.01	.50	.02
O5 They were integrated into the course	.00	-.13	.14	.19	-.02	.28	.01	.21	.44	.05
X WORKLOAD/DIFFICULTY										
M32 Course difficulty was: (Easy-Hard)	-.07	-.11	.03	-.03	.03	.02	.05	-.04	-.02	.79
M33 Course workload was: (Light-Heavy)	.12	-.10	-.02	.11	-.05	.09	-.04	-.02	.04	.70
F4 Students had to work hard	.05	.10	-.02	-.02	-.02	-.07	-.08	.08	.10	.77
F5 Course required a lot of time	-.01	.11	-.02	.02	.05	-.02	.05	-.02	.03	.85
M30 OVERALL COURSE RATING	.50	.07	.23	.06	.00	.10	.15	.12	.05	.01
M31 OVERALL INSTRUCTOR RATING	.26	.28	.43	.03	.07	.14	.14	.03	.15	-.03

NOTE: The oblique factor analysis was performed with the commercially available SPSS routine (Nie, et. al., 1975). Factor loadings in the boxes are the loadings for items designed to measure each factor.